

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2004-029943

(43)Date of publication of application : 29.01.2004

(51)Int.Cl. G06F 17/30
G06F 13/00

(21)Application number : 2002-181725

(71)Applicant : RECRUIT CO LTD

(22)Date of filing : 21.06.2002

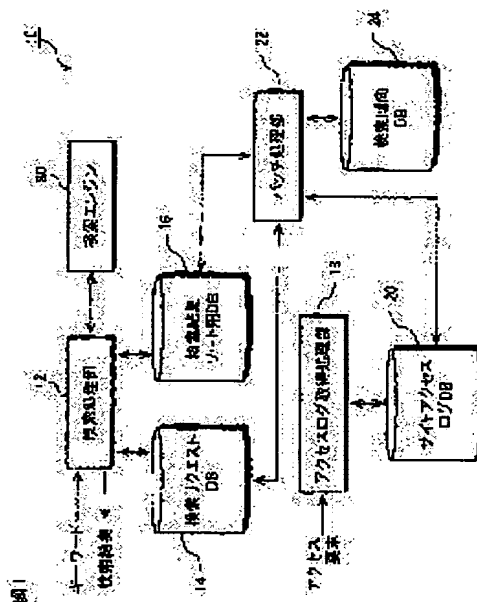
(72)Inventor : TAIRA MASANOBU

(54) RETRIEVAL SUPPORT METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To present retrieval results to allow a user to more easily find out his or her desired information.

SOLUTION: At receiving a retrieval keyword, a retrieval processing part 12 and a retrieval engine 30 acquire the information of one or more URL for specifying a site related with the retrieval keyword. The retrieval processing part 12 refers to a DB 16 for sorting retrieval results in which retrieval keywords, URL, and indexes indicating the extents of the estimated achievement of the goal in the sites of the URL are stored, and sorts the acquired URL in the order of the indexes, and generates retrieval result contents including the list for introducing the sites according to the order of sort of the URL, and presents this to a client machine.



LEGAL STATUS

[Date of request for examination]

17.06.2005

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2004-29943

(P2004-29943A)

(43) 公開日 平成16年1月29日(2004.1.29)

(51) Int. Cl.⁷

G06F 17/30

G06F 13/00

F 1

G06F 17/30 340B

G06F 17/30 110F

G06F 17/30 380E

G06F 13/00 540E

テーマコード (参考)

5B075

審査請求 未請求 請求項の数 7 O L (全 14 頁)

(21) 出願番号 特願2002-181725 (P2002-181725)
(22) 出願日 平成14年6月21日 (2002.6.21)(71) 出願人 000139012
株式会社リクルート
東京都中央区銀座8丁目4番17号
(74) 代理人 100099715
弁理士 吉田 聡
(72) 発明者 平 昌展
東京都中央区銀座8丁目4番17号 株式
会社リクルート内
Fターム(参考) 5B075 KK07 NK02 NK44 NR05 NR15
PP22 PQ02 PQ27 PQ46 PQ72
PQ74 PR01 PR04 PR06 PR08

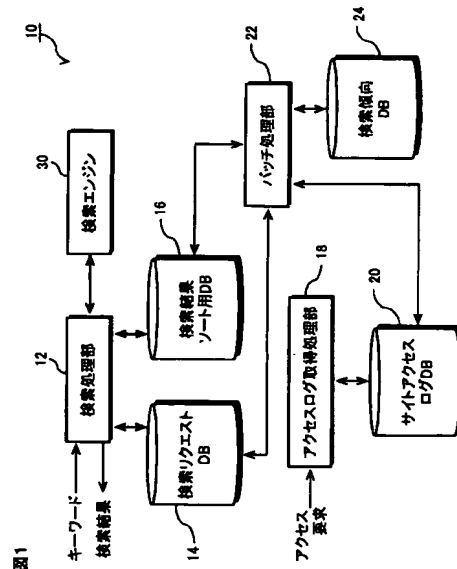
(54) 【発明の名称】 検索支援方法

(57) 【要約】

【課題】 ユーザが所望の情報を、より容易に見出すことができるような検索結果を提示する。

【解決手段】 検索処理部12および検索エンジン30は、検索キーワードを受理すると、当該検索キーワードに関連するサイトを特定する1以上のURLの情報を取得する。検索処理部12は、検索キーワード、URL、および、当該URLのサイトにて目的が達成されたと推定されることの度合いを示す指標を記憶した検索結果ソート用DB16を参照して、取得されたURLを、前記指標の順にソートし、URLのソート順にしたがって、サイトを紹介するリストを含む検索結果コンテンツを生成して、クライアントマシンに提示する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

1 以上の検索キーワードを受理して、当該検索キーワードに関連するサイトの情報のリストを生成して、ユーザの操作するクライアントマシンに提示する検索支援方法であって、検索キーワードを受理し、当該検索キーワードに関連するサイトを特定する 1 以上の URL の情報を取得するステップと、

前記検索キーワード、URL、および、当該 URL のサイトにて目的が達成されたと推定されることの度合いを示す指標を記憶したソート用データベースを参照して、前記取得された URL を、前記指標の順にソートするステップと、

前記 URL のソート順にしたがって、前記サイトを紹介するリストを含む検索結果コンテンツを生成し、クライアントマシンに提示するステップと、

前記クライアントマシンにおける前記コンテンツ中、特定のサイトのリンク指定を受理して、前記サイトの URL を取得し、アクセスログデータベースに記憶するとともに、リダイレクトにより、前記サイトのコンテンツのクライアントマシンへの配信を求めるステップと、

前記アクセスログデータベースを参照して、前記取得された URL のうち、ある検索キーワードを用いた検索においてユーザが目的を達成したと推定されるサイトの URL を特定するステップと、

前記検索キーワードに関して、目的を達成したと推定されるサイトの URL に関する指標を再計算し、前記ソート用データベースを更新するステップとを備えたことを特徴とする検索支援方法。

【請求項 2】

さらに、ユーザのアクセスごとに検索 ID を発行するステップを備え、

前記検索結果コンテンツに基づくリンクの指定が維持される限り、同一の検索 ID を利用するとともに、

前記ソート用データベースを更新するステップが、各検索 ID に関して指標を再計算するステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 3】

さらに、ある検索において、リンク指定の回数を、前記アクセスログデータベースに記憶するステップを備え、

ユーザの検索傾向を示すアクセス回数上限値を記憶した検索傾向データベースを参照して、前記リンク指定の回数が、前記アクセス回数上限値を超えた場合に、当該検索における URL を、前記指標の再計算には考慮しないことを特徴とする請求項 2 に記載の方法。

【請求項 4】

前記アクセス回数上限値が、当該ユーザによる検索ごとのリンク指定の回数の情報信頼限界に対応することを特徴とする請求項 3 に記載の方法。

【請求項 5】

前記目的を達成したと推定されるサイトの URL を特定するステップにおいて、ある検索において、あるリンク指定から、一定の時間、他のリンク指定がなされていない状態である場合に、当該リンク指定にかかる URL が、前記目的を達成したと推定されるものとして特定されることを特徴とする請求項 1 ないし 4 の何れか一項に記載の方法。

【請求項 6】

前記指標が、ある検索において目的を達成したと推定されるサイトごとに加点されるポイントであることを特徴とする請求項 1 ないし 5 の何れか一項に記載の方法。

【請求項 7】

さらに、取得された URL の情報のうち、前記ソート用データベース中に存在しないものについて、当該 URL のサイトに含まれる検索キーワードの出現頻度や当該 URL のサイトの支持率を含む他の論理に基づく順序で、当該リストを含むコンテンツを生成するステップを備えたことを特徴とする請求項 1 ないし 6 の何れか一項に記載の方法。

【発明の詳細な説明】

10

20

30

40

50

【 0 0 0 1 】

【 発明が属する技術分野 】

本発明は、検索サイトにおける検索結果を提示し、ユーザによるサイト検索を支援する手法に関する。

【 0 0 0 2 】

【 従来の技術 】

インターネットの普及により、ユーザは、種々のサイトにアクセスし、その場でリアルタイムに、ユーザが求める情報を取得できるようになっている。必要な情報を含むサイトを見出すために、検索サイトの利用が最も有用である。検索サイトの使い方には、ユーザにより入力された検索キーワードに基づき、コンテンツ中に当該検索キーワードを含むようなサイトを見出すようなキーワード検索と、カテゴリを絞りつつ目的となるサイトを見出すようなディレクトリ検索とが知られている。

10

【 0 0 0 3 】

たとえば、キーワード検索の場合、ある検索キーワードを入力すると、サーバに備えられた検索エンジンが、当該検索キーワードを含むサイトを特定し、そのURLを、当該サイトの要約や、検索キーワードを含む文章などとともに、検索結果としてリスト（検索結果コンテンツ）し、クライアントマシンに返すようになっている。

【 0 0 0 4 】

ユーザは、それぞれのサイトの要約などを参照し、所望の情報を取得できると期待されるサイト等のリンクを指定し、そのコンテンツを取得する。ユーザは、検索結果コンテンツ中の要約などの参照、サイトへのリンク指定を繰り返して、何れかのサイトに、ユーザが所望の情報が見出された段階で、そのサイトに留まることになる。

20

【 0 0 0 5 】

【 発明が解決しようとする課題 】

ユーザが上記手順を繰り返して、余計な時間を費やすことなく、所望の情報が含まれるサイトに到達するためには、検索結果コンテンツ中に、ユーザが求める情報を含むサイトのリストが、より上位に位置し、或いは、より目立つ状態であるのが望ましい。

【 0 0 0 6 】

従来の検索エンジンにおいては、たとえば、サイト中に、どのくらい多数の検索キーワードが含まれているかに応じて、リストの配置を決定し、または、サイトの支持率、つまり、そのサイトがどのくらい他のサイトからリンクされているかにしたがって、リストの配置を決定している。たとえば、前者では、検索キーワードの出現頻度が多いものは、当該検索キーワードとの関連が深く、適切なサイトであるという仮定に基づいている。また、後者では、他のサイトからのリンクがそのサイトを評価している指標になるという仮定に基づいている。

30

【 0 0 0 7 】

また、知識ベースを利用し、ユーザが入力した検索キーワードの上位概念に対応する語句を見出し、その語句を含むサイトのリストを生成し、ユーザの検索キーワードの選択を実質的に補助するような検索サイトも知られている。

しかしながら、従来の検索サイトにおいては、主として出現頻度や支持率に基づく仮定により、リストの順位を決定しているため、実際に、ユーザには、どのサイトに、有用な情報が含まれていたかを知ることができなかった。このため、たとえば、下位にリストされているにもかかわらず、数多くのユーザがそこから所望の情報を得ているようなサイトを、ユーザに適切に紹介することができず、このため、ユーザに余分な検索時間や検索の労力を負わせることになるという問題点があった。

40

【 0 0 0 8 】

本発明は、ユーザが所望の情報を、より容易に見出すことができるような検索結果を提示することができる検索支援方法を提供することを目的とする。

【 0 0 0 9 】

【 課題を解決するための手段 】

50

本発明の目的は、1以上の検索キーワードを受理して、当該検索キーワードに関連するサイトの情報のリストを生成して、ユーザの操作するクライアントマシンに提示する検索支援方法であって、検索キーワードを受理し、当該検索キーワードに関連するサイトを特定する1以上のURLの情報を取得するステップと、前記検索キーワード、URL、および、当該URLのサイトにて目的が達成されたと推定されることの度合いを示す指標を記憶したソート用データベースを参照して、前記取得されたURLを、前記指標の順にソートするステップと、前記URLのソート順にしたがって、前記サイトを紹介するリストを含む検索結果コンテンツを生成し、クライアントマシンに提示するステップと、前記クライアントマシンにおける前記コンテンツ中、特定のサイトのリンク指定を受理して、前記サイトのURLを取得し、アクセスログデータベースに記憶するとともに、リダイレクトにより、前記サイトのコンテンツのクライアントマシンへの配信を求めるステップと、前記アクセスログデータベースを参照して、前記取得されたURLのうち、ある検索キーワードを用いた検索においてユーザが目的を達成したと推定されるサイトのURLを特定するステップと、前記検索キーワードに関して、目的を達成したと推定されるサイトのURLに関する指標を再計算し、前記ソート用データベースを更新するステップとを備えたことを特徴とする検索支援方法により達成される。

10

【0010】

本発明によれば、検索キーワードに関連するサイトが、ユーザが目的を達成したと推定される、つまり、所望の情報が取得されたと推定される度合いを示す指標が大きいものの順にリストされた形式の検索結果コンテンツがユーザに配信される。たとえば、上記指標の大きいものから順に、サイトの紹介をリストしても良いし、或いは、指標の大きいものをより目立つような表示を施しても良い。指標の大きなサイトを閲覧することで、ユーザは所望の情報を短時間で取得できる可能性を高めることが可能となる。

20

【0011】

上記ソート用データベースを更新するステップは、たとえば、夜間などのバッチ処理で実行され、その一方、他のステップは、ユーザが検索を実行している際に、リアルタイム処理で実行され得る。

【0012】

好ましい実施態様においては、さらに、ユーザのアクセスごとに検索IDを発行するステップを備え、前記検索結果コンテンツに基づくリンクの指定が維持される限り、同一の検索IDを利用するとともに、前記ソート用データベースを更新するステップが、各検索IDに関して指標を再計算するステップを含む。

30

【0013】

また、好ましい実施態様においては、さらに、ある検索において、リンク指定の回数を、前記アクセスログデータベースに記憶するステップを備え、ユーザの検索傾向を示すアクセス回数上限値を記憶した検索傾向データベースを参照して、前記リンク指定の回数が、前記アクセス回数上限値を超えた場合に、当該検索におけるURLを、前記指標の再計算には考慮しないように構成されている。この検索傾向データベースも、アクセスログデータベースを参照して、バッチ処理にて更新されうる。これにより、上記アクセス回数上限値もダイナミックに更新され得る。

40

【0014】

これは、ユーザによる検索の癖（傾向）を把握し、その傾向から逸脱しているような検索を無視している。これにより、指標の算出精度を高めることが可能となる。前記アクセス回数上限値として、当該ユーザによる検索ごとのリンク指定の回数の情報信頼限界を利用することができる。なお、検索キーワードをカテゴライズし、カテゴリごとにアクセス回数上限値を別途設定しても良い。

別の好ましい実施態様において、前記目的を達成したと推定されるサイトのURLを特定するステップにおいて、ある検索において、あるリンク指定から、一定の時間、他のリンク指定がなされていない状態である場合に、当該リンク指定にかかるURLが、前記目的を達成したと推定されるものとして特定される。

50

【0015】

また、好ましい実施態様においては、前記指標が、ある検索において目的を達成したと推定されるサイトごとに加点されるポイントである。

より好ましい実施態様においては、さらに、取得されたURLの情報のうち、前記ソート用データベース中に存在しないものについて、当該URLのサイトに含まれる検索キーワードの出現頻度や当該URLのサイトの支持率を含む他の論理に基づく順序で、当該リストを含むコンテンツを生成するステップを備えている。

【0016】

【発明の実施の形態】

以下、添付図面を参照して、本発明の実施の形態につき説明を加える。図1は、本発明の実施の形態にかかる検索支援サーバ（以下、「サーバ」と称する。）の構成を示すブロックダイアグラムである。図1に示すサーバは、ネットワーク、たとえば、インターネットに接続され、クライアントマシンから与えられるキーワードを受理して、関連するサイトを検索し、サイトの情報を含む検索結果をコンテンツとしてクライアントマシンに返すようになっている。なお、本明細書において、サイトとは、ウェブページの集合体、当該集合体のトップページ、集合体を構成するそれぞれのウェブページを指すものとする。

【0017】

図1に示すように、本実施の形態にかかるサーバ10は、ユーザが入力したキーワードを、インターネットを介して受理し、サイト検索に関する種々の処理を実行する検索処理部12と、アクセスしてくるユーザが、どのようなキーワードを利用したかを示す情報を記憶した検索リクエストデータベース（DB）14と、検索結果を提示するために利用される、キーワードおよびURLの組のそれぞれについて付与されたポイントを記憶した検索結果ソート用DB16と、ユーザがクライアントマシンを操作することにより、サーバに与えられる、他のサイトへのアクセス要求を受理し、ユーザによるアクセスログを取得するアクセスログ取得処理部18と、ユーザによるアクセスログを記憶するサイトアクセスログDB20と、ユーザによるアクセスログやユーザが利用した検索キーワードなど、種々のアクセス履歴にしたがって、適切な検索結果の提示のために必要なデータを用意する処理を実行するパッチ処理部22と、ユーザによる検索傾向を記憶した検索傾向DB24とを有している。

【0018】

また、検索処理部12には、実際に、検索キーワードを参照して、当該検索キーワードに関連するサイトのURLを取得する検索エンジン30が接続されている。検索エンジン30は、サーバ10に含まれていても良いし、サーバ10と別体であっても良い。

【0019】

このように構成されたサーバ10を利用した検索およびサイトへのアクセスにつき、図2を参照して説明を加える。ユーザがクライアントマシンを操作して、検索キーワードを入力すると、当該検索キーワードがサーバに伝達される（ステップ201）。サーバ10においては、後述する検索処理が実行され（ステップ202）、検索結果がリストされたコンテンツ（結果コンテンツ）が、クライアントマシンに返される（ステップ203）。

【0020】

結果コンテンツには、サイト名や、検索キーワードを含むサイト中の文字列などが示される。本実施の形態においては、ユーザが、サイト名をクリックしても、直接、そのサイトのURLにアクセス要求が伝達されるのではなく、いったん、サーバ10を介して、当該サイトへのアクセス要求が伝達されるようになっている。このため、サイト指定（アクセス要求）は、サーバ10に伝えられ（ステップ204）、サーバ10において、ユーザがアクセスを要求したURLを含むアクセスログが取得された後（ステップ205）、いわゆるリダイレクト処理により、サーバ10から、指定された他のサイトへのアクセス要求が発せられる（ステップ206）。他のサイトからのコンテンツは、クライアントマシンに配信され（ステップ208）、クライアントマシンのブラウザにより、その表示装置の画面上にコンテンツが表示される。

10

20

30

40

50

【0021】

たとえば、ユーザが入力装置を操作して、ブラウザのバックボタン（「戻る」ボタン）をオンすれば、再度、検索結果がリストされた結果コンテンツが表示される。その状態で、ユーザが、入力装置を操作してサイトを指定すると（ステップ209）、再度、アクセスログ取得処理（ステップ210）、他のサイトへのアクセス要求（ステップ211）が実行されて、指定されたコンテンツがクライアントマシンに配信される（ステップ212）。

このように、結果コンテンツにリストされたサイト名などのリンクを指定することにより、サーバ10は、ユーザが閲覧しようとしたサイトのURLを含むアクセスログを取得し、これを蓄積することができる。

10

【0022】

次に、本実施の形態にかかる検索処理をより詳細に説明する。図3は、本実施の形態にかかる検索処理を示すフローチャートである。検索処理部12は、クライアントマシンからの検索キーワードの受理に応答して、まず、当該クライアントマシンを利用してユーザにより既にサーバにアクセスされ、ユーザに、ユーザIDを含むCookie（クッキー）が送信されているか否かを判断する（ステップ301）。ユーザIDが存在しない場合には、当該クライアントマシンを操作しているユーザに対してユーザIDを発番する（ステップ302）。

【0023】

次いで、今回の検索行為を一意的に特定するための検索IDが発番される（ステップ303）。この検索IDは、クライアントマシンから、ある検索キーワード（単一の検索キーワード、或いは、複数の検索キーワードの組み合わせ、以後、本明細書において同様である。）がサーバ10に与えられ、当該検索キーワードに基づく検索結果に基づいて、クライアントマシンから、サイト指定がサーバ10に与えられる間だけ維持される。つまり、検索IDと検索キーワードとは、一意的に関連付けされている。したがって、ユーザが、新たな検索キーワードを入力し、或いは、ある検索キーワードに、さらに他の検索キーワードを付加した場合には、異なる検索IDが付与される。

20

【0024】

次いで、ユーザが入力した検索キーワードが、それぞれ正規化され、その後、ソートされる（ステップ304、305）。ステップ304においては、正規化により、ユーザごとの入力によるゆらぎ（たとえば、全角入力／半角入力、英大文字／小文字など）が、一定の基準にて統一された表現とされる。また、ユーザが、複数の検索キーワードをスペース等で区切って入力する場合がある。この場合には、入力された複数の検索キーワードの「AND」検索となる。このように入力キーワードが複数である場合に、ステップ305のソート処理によりユーザが入力した語順にかかわらず、同一の検索キーワードであれば、同じ検索結果が得られるようにしている。

30

【0025】

次いで、ログデータを保持するために、正規化およびソート処理が施された検索キーワードのレコードが、検索リクエストDB14に追加される。図4（a）は、検索リクエストDB14中に記憶されるデータの例を示す図である。図4（a）に示すように、検索リクエストDB14においては、検索IDごとに、その検索に利用された検索キーワードが関連付けられて記憶されている。また、後述する処理により、当該検索キーワードによる検索結果を利用して、ユーザが何回、リンクを指定したかを示す検索回数が、検索IDごとに算出されて記憶される。初期的には、検索回数は「0（ゼロ）」に設定される。また、前述したように、検索キーワードの入力や追加ごとに検索IDが発番されるため、あるユーザID（たとえば、ユーザID=GAW0023514）に対して、複数の検索IDが存在し得る。

40

【0026】

次いで、検索エンジン30が、正規化およびソートされた検索キーワードを利用して、当該検索キーワードに関連したサイトを検索する（ステップ307）。検索エンジン30に

50

よる検索結果を示すリストには、一定の順序でサイトのURLが挙げられている。この順序として、たとえば、サイトがどの程度支持されているか（そのサイトへのリンクがどのくらい存在しているか）、キーワードの出現頻度などが利用されている。

【0027】

検索エンジン30からの検索結果を受理すると、検索処理部12は、検索結果ソート用DB16を参照して、今回の検索IDにかかる検索キーワードが、検索結果ソート用DB16中に存在するか否かを判断する（ステップ308、309）。図4（b）は、検索結果ソート用DB16のデータの例を示す図である。検索ソート用DB16においては、検索キーワードおよびURLの組み合わせと、ポイントとが関連付けられている。ここで、ポイントは、後述するポイント付与処理（図10参照）により生成された、ある検索キーワードにて見出されるURLが、どの程度有用であることを示す指標である。

10

【0028】

ステップ309でイエス（Yes）と判断された場合には、検索処理部12は、受理した検索結果に含まれるURLを、ポイントの高い順にソートし（ステップ310）、かつ、検索結果ソート用DB16の検索キーワードに関連付けられていないURLに関しては、受理した検索結果の順序で、URLのリストを生成する（ステップ311）。その一方、ステップ309でノー（No）と判断された場合には、受理した検索結果の順序で、URLのリストを生成する（ステップ311）。このようにして、検索結果コンテンツが生成される。

【0029】

前述したように、ユーザのサイト指定に応答して、サーバ10においてアクセス取得処理が実行されて（ステップ205）、アクセスログが取得されるとともに、リダイレクト処理により、ユーザが指定したサイトから、クライアントマシンにコンテンツが配信される（ステップ206、207）。図5は、アクセスログ取得処理をより詳細に示すフローチャートである。

20

【0030】

アクセスログ取得処理部18は、受理した情報に基づき、サイトアクセスログDB20に、当該サイト指定にかかるレコードを追加する（ステップ501）。図8（a）は、サイトアクセスログDB20中に記憶されたデータの例を示す図である。図8（a）に示すように、サイトアクセスログDB20においては、検索ID、連番、URLおよびアクセスログ日時の組からなるレコードが記憶される。本実施の形態においては、図2に示したように、ある検索結果コンテンツが提示されている状態から、サイト指定、アクセスログ取得処理、コンテンツ配信が繰り返され得る（ステップ204～207、および、ステップ209～212参照）。そこで、アクセスログ取得処理部18は、ある検索IDに関して、ユーザがサイト指定をするたびに、インクリメントされた連番を採番して、これを含むレコードを生成し、サイトアクセスログDB20に追加する。

30

【0031】

次いで、リダイレクト処理（ステップ502）により、ユーザが指定したサイトに対して、クライアントマシンに対するコンテンツの配信を要求する。これに応じて、指定されたサイトからクライアントマシンにコンテンツが配信され、クライアントマシンのブラウザによりこれが表示される。

40

【0032】

クライアントマシンからのアクセスに応答して、サーバ10は、上述した処理を実行する。その一方、サーバ10は、夜間など所定のタイミングで、バッチ処理を実行し、サイトアクセスログDB20や検索リクエストDB14からのデータを取得して、検索傾向DB24や検索結果ソート用DB16を更新している。図6は、バッチ処理を示すフローチャートである。バッチ処理においては、検索ID抽出処理（ステップ601）および検索傾向算出処理（ステップ602）が実行される。

【0033】

検索ID抽出処理においては、後述する前回抽出対象日時ファイル611に基づいて、処

50

理対象となるレコードを、サイトアクセスログDB20から抽出し、今回抽出サイトアクセスログファイル612を生成する。また、検索傾向算出処理においては、検索リクエストDB14から、各ユーザに関するレコードを取り出して、当該ユーザの検索の癖（傾向）を示す指標を算出して、検索傾向DB24を更新するとともに、検索結果ソート用DB16を更新する。

【0034】

これらにつき、図7、図8および図10を参照してより詳細に説明を加える。図7に示すように、検索ID抽出処理においては、まず、バッチ処理部22は、前回抽出対象日時ファイル611から、前回抽出対象となった、サイトアクセスログDB20中のレコードのアクセス日時の上限値を取得する（ステップ701）。ここで、前回抽出対象日時とは、前回バッチ処理において処理対象とすべきアクセス日時の上限値（前回バッチ処理開始日時－6時間）を言う。本実施の形態においては、検索IDにて特定されるある検索行為において、最も大きい連番を有するアクセスのアクセス日時が、前回抽出対象日時より大きく、かつ、今回バッチ処理開始日時からすでに6時間を経過した場合に、抽出対象となっている。したがって、アクセスログ抽出処理（ステップ702）においては、検索条件として、「前回抽出対象日時＜ある検索IDに関して最終連番を有するレコードにおけるアクセス日時≤現在時刻－6時間」を利用し、当該検索条件を満たす検索IDをもつレコードが抽出される。

10

【0035】

たとえば、図11に示すように、ある検索ID（たとえば、検索ID＝GAW00235140000001）に関して、連番4ないし連番6がそれぞれ付与されたアクセス（符号1104～1106参照）があったと考える。この場合には、前回のバッチ処理開始日時（符号1110参照）の「現在時刻－6時間」、つまり、前回抽出対象日時（符号1111参照）以後に、連番5のアクセスが存在している。したがって、前回のバッチ処理時においては、当該検索IDを有するアクセスは、抽出対象とはならない。その一方、当該検索IDの連番6のアクセス（符号1106参照）は、今回のバッチ処理時において検索条件を満たすため、当該検索IDを有するアクセスは、すべて抽出対象となる。同様に、他の検索ID（たとえば、検索ID＝GAW00235140000002）に関して、連番1が付与されたアクセス（符号1121参照）が、「現在時刻（今回バッチ処理開始日時）－6時間」より以前に存在するため、この検索IDを有するレコードは抽出対象たり得る。

20

30

【0036】

このように、現在時刻（今回バッチ処理開始日時）から一定の時間（本実施の形態では6時間）以前に、最終の連番が存在する場合のみ、検索IDにかかるレコードを抽出する理由につき説明を加える。図2を参照して説明したように、本実施の形態においては、ユーザが、ある検索キーワードを入力し、その検索結果を受けた後、当該ユーザが、いくつかのサイトを閲覧し、閲覧したサイトに、実際にユーザが求めている情報が含まれている場合には、そのサイトに留まること、つまり、他のサイトを閲覧しないと想像される。そこで、本実施の形態においては、サイトに留まっていると判断するために、一定の時間が設定されている。

40

【0037】

このようにして、今回抽出サイトのアクセスログのファイル612が作成されると、検索回数更新処理が実行される（ステップ703）。検索回数更新処理703においては、今回抽出サイトアクセスログファイル612中のレコード数を、検索IDごとにカウントする。たとえば、図11の例では、今回のバッチ処理において、検索ID＝GAW00235140000001のカウント値は、当該検索IDの連番が6までであるため、6となる。その一方、検索ID＝00235140000002のカウント値は1となる。このカウント値は、検索リクエストDB14中の、対応する検索IDの検索回数として記憶される。

【0038】

次に、図9を参照して、検索傾向算出処理につき説明を加える。検索傾向算出処理におい

50

ては、まず、検索IDをキーとして、検索リクエストDB14中のデータがソートされる（ステップ901）。これは、ソート後の検索リクエストDB中のデータがなくなるまで繰り返される。次いで、ユーザID単位で、当該ユーザIDに関する全検索IDのもつ検索回数の平均値と標準偏差が算出される（ステップ902）。その後、ユーザIDごとのアクセス回数上限値が算出され、これが検索傾向DB24に、ユーザIDと関連付けて記憶される（ステップ903）。

【0039】

本実施の形態においては、ユーザ単位で、単一の検索キーワードを利用した検索回数は正規分布すると考え、信頼係数95パーセントとして、情報信頼限界（ $=1.96 \times \text{標準偏差} + \text{平均値}$ ）を算出し、これを当該ユーザのアクセス回数上限値とした。本実施の形態においては、このアクセス回数上限値が、ユーザの検索の癖（傾向）を示す指標となる。

10

【0040】

次いで、全ユーザのアクセス回数の平均値が算出され（ステップ905）、これも、検索傾向DB24に記憶される。なお、図8（b）において、ユーザID=00000000000に対応したアクセス回数上限値が、ステップ905で算出された平均値である。この平均値は、サーバ10に初めてアクセスしたため、そのアクセス回数上限値が設定されていないようなユーザのためのデフォルト値として利用することができる。

【0041】

検索傾向DB24の更新が終了すると、検索ID抽出処理が実行される（ステップ905）。ここでは、検索傾向DBを参照しつつ、ユーザIDごとに、当該ユーザのアクセス回数上限値以内の検索回数であるような、検索IDおよび検索キーワードの組を、検索リクエストDB14から抽出し、検索IDおよび検索キーワードの組からなるファイル911が生成される。このファイル911を利用して、ポイント付与処理が実行される（ステップ906）。

20

【0042】

図10は、ポイント付与処理をより詳細に示すフローチャートである。まず、ファイル911について、検索キーワードをキーとしてソートし、検索キーワードおよび検索IDの組からなるファイル1011が生成される。これは、ソートし直したファイルのデータがなくなるまで継続される（ステップ1011、1002）。次いで、検索IDごとに、サイトアクセスログDB20を参照して、当該検索IDが最終的にアクセスしたログ（URL）が取得される（ステップ1003）。これにより、検索キーワード、検索IDおよびURLの組からなるファイル1012が生成される。このファイル1012を参照して、当該ファイル1012中のある検索キーワードおよびURLが、検索結果ソート用DB16中に存在するか否かが判断される（ステップ1004、1005）。すでに上記組み合わせが存在する場合には（ステップ1005でイエス（Yes））、検索結果ソート用DB16中、対応するレコードにおけるポイントがインクリメントされる（ステップ1006）。その一方、ステップ1005でノー（No）と判断された場合には、上記検索キーワードおよびURLの組み合わせを含むレコードが、検索結果ソート用DB16中に生成され、かつ、その対応するポイントが1に設定される（ステップ1007）。

30

【0043】

このようにして、各種DBのデータを更新するバッチ処理が終了する。再度、図3を参照して、これら更新されたデータがどのように利用されるかにつき簡単に説明する。ステップ307において検索エンジン30から検索結果が検索処理部12に返される。検索処理部12は、検索結果ソート用DB16中、検索キーワードに関連付けられたURLであって、検索結果と一致するものを見出して、それらを、ポイントの高いものから順に並べてリストする。これにより、ポイントの高いもの、つまり、他のユーザによる検索結果の蓄積から、有用と考えられる度合いの大きいものから順に、サイトの情報がリストされることになる。これにより、ユーザは、リストの上位に位置しているものから順次、サイトを指定して、その内容を閲覧し、ユーザが求める情報が含まれるか否かを判断すれば良い。この場合にも、あまりサイト指定を繰り返すことなく、所望の情報が含まれるサイトを見

40

50

出すことができる可能性が高いことが期待される。

【0044】

このように、本実施の形態によれば、過去にユーザが検索を実行し、どのサイトで目的の情報を見出されたかを推定し、その推定から、検索キーワードおよびサイト（URL）にポイントが付与し、このポイントを利用して、後の検索におけるサイトの提示順序を決定している。したがって、たとえば、上位にリストされたサイトを指定すれば、ユーザが所望の情報を見出すことができる確率が高まり、ユーザが、情報を見出すために何度もサイトを行き来するような手間を減じることが可能となる。

【0045】

また、本実施の形態によれば、検索傾向DBに、各ユーザの検索の癖を示す指標である検索回数を利用し、この検索回数を超えてサイト指定を繰り返している場合には、この検索は信頼できないものとして考慮しないように構成している。これにより、上記ポイントの信頼性を高めることが可能となる。また、上記検索回数を、バッチ処理により更新することにより、データの信頼性をより高めている。

【0046】

本発明は、以上の実施の形態に限定されることなく、特許請求の範囲に記載された発明の範囲内で、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

たとえば、単一のユーザIDをもつユーザ（単一のユーザ）が、異なる検索IDを付与された検索において、同一の検索キーワードを利用する場合も考えられる。この場合に、本実施の形態のように、異なる検索IDであれば、異なるものとして処理しても良い。或いは、所定の時間内或いは所定のセッション内に、同一の検索キーワードが利用された場合には、アクセスログをマージして、最終的なURLのみを保持するように構成しても良い。

【0047】

また、前記実施の形態において、サイトアクセスログDBには、ある検索IDに関してサイト指定されたURLの全てが、一意的な連番とともに記憶されているが、これに限定されるものではなく、最新の連番および最新のURLのみを記憶するような構成としても良い。

【0048】

さらに、前記実施の形態においては、検索結果ソート用DBにおいて、単一の検索キーワード（たとえば、「C++」、「C」など）、複数の検索キーワードからなる複合語（たとえば、「C++およびC」など）のそれぞれについて、URLおよびポイントが保持されている。しかしながら、これに限定されるものではなく、単一の検索キーワードに関して、URLおよびポイントを保持しても良い。この場合には、ポイント付与処理において、複数の検索キーワードのそれぞれに関して、URLにポイントを付与すればよい。また、検索処理においても、複数の検索キーワードのそれぞれに関連付けられるURLを見出し、それらのポイントの加算値、平均値などを利用して、複数の検索キーワードに関するURLのポイントを求め、そのポイントの上位から、URLをリストしても良い。

【0049】

また、アクセス回数上限値を、キーワードのカテゴリに応じて、別途設定しても良い。たとえば、同じユーザであっても、ビジネスユースであれば、所望の情報を得るために検索回数が上昇し、その一方、趣味など個人的な利用であれば、検索回数が少なくなる傾向がある。そこで、検索キーワードを、主としてビジネスにて利用されるものと、主として趣味で利用されるものにカテゴリ化し、それぞれに応じたアクセス回数上限値を設定することもできる。

【0050】

【発明の効果】

本発明によれば、ユーザが所望の情報を、より容易に見出すことができるような検索結果を提示することができる検索支援方法を提供することが可能となる。

【図面の簡単な説明】

【図 1】図 1 は、本発明の実施の形態にかかる検索支援サーバの構成を示すブロックダイヤグラムである。

【図 2】図 2 は、本実施の形態にかかるサーバを利用した検索およびサイトへのアクセスを示すフローチャートである。

【図 3】図 3 は、本実施の形態にかかる検索処理をより詳細に示すフローチャートである。

【図 4】図 4 は、検索リクエスト DB 中に記憶されるデータ、および、検索結果ソート用 DB 中に記憶されるデータの例を示す図である。

【図 5】図 5 は、本実施の形態にかかるアクセスログ取得処理をより詳細に示すフローチャートである。

10

【図 6】図 6 は、本実施の形態にかかるパッチ処理の概略を示すフローチャートである。

【図 7】図 7 は、本実施の形態にかかる検索 ID 抽出処理をより詳細に示すフローチャートである。

【図 8】図 8 は、本実施の形態にかかるアクセスログ DB 中に記憶されたデータ、および、検索傾向 DB 中に記憶されたデータの例を示す図である。

【図 9】図 9 は、本実施の形態にかかる検索傾向算出処理をより詳細に示すフローチャートである。

【図 10】図 10 は、本実施の形態にかかるポイント付与処理をより詳細に示すフローチャートである。

20

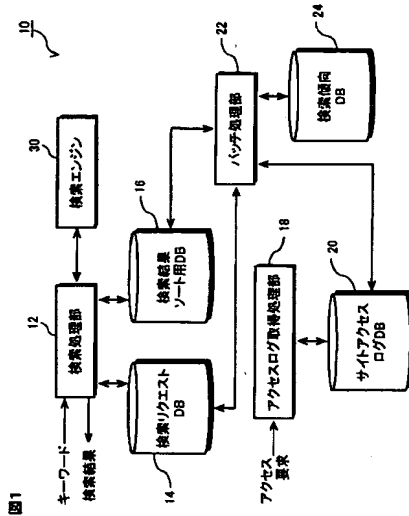
【図 11】図 11 は、本実施の形態にかかるアクセスログ抽出処理を説明するための図である。

【符号の説明】

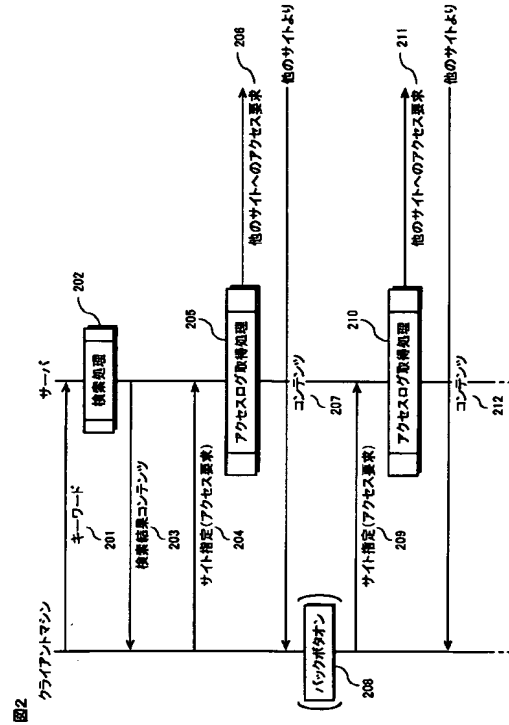
- 10 検索支援サーバ
- 12 検索処理部
- 14 検索リクエスト DB
- 16 検索結果ソート用 DB
- 18 アクセスログ取得処理部
- 20 サイトアクセスログ DB
- 22 パッチ処理部
- 24 検索傾向 DB
- 30 検索エンジン

30

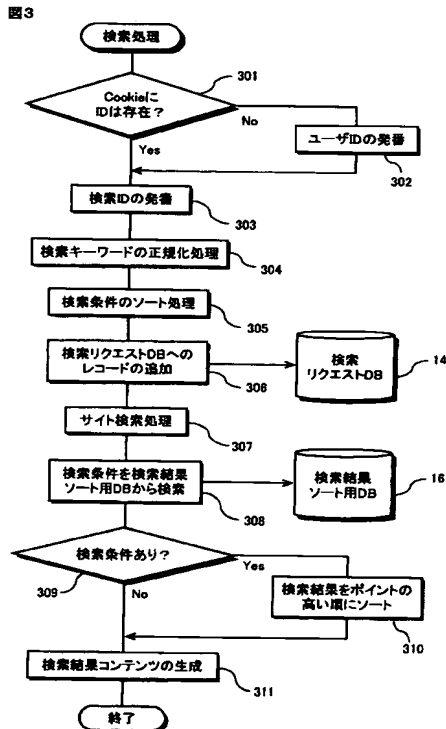
【図 1】



【図 2】



【図 3】



【図 4】

34

(a)

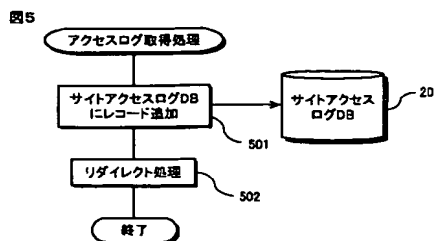
ユーザID	検索ID	検索キーワード	検索回数
GAW0023514	GAW0023514000001	C**	3
	GAW0023514000002	C	2
	GAW0023514000003	JAVA⊙	4
	GAW0023514000004	Perl	4
	GAW0023514000005	Apache module	3
	GAW0023514000006	BIND セキュリティホール	3
	GAW0023514000007	C**	4
	GAW0023514000008	ハックアフォーバーフロー	5
	GAW0023514000009	Grid	3
	GAW0023514000010	シェルコマンド	4
JPS0086754	JPS0086754000001	映画	4
	JPS0086754000002	音楽	6
	

(b)

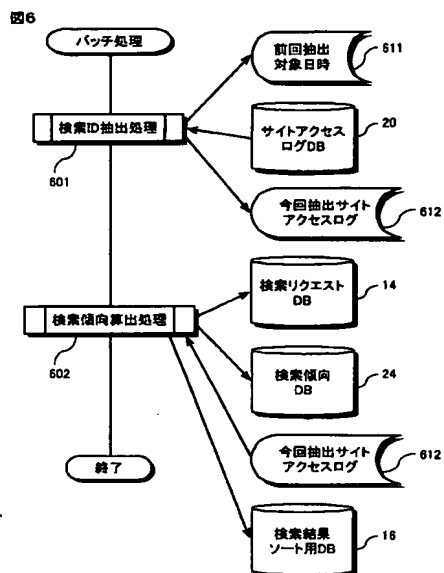
検索キーワード	URL	ポイント
C**	aaa.bbb.org/	1329
C	www.ccc.com/ba/C*.htm	2545
JAVA⊙	www.ddd.com/	4345
C**
.....

図 4

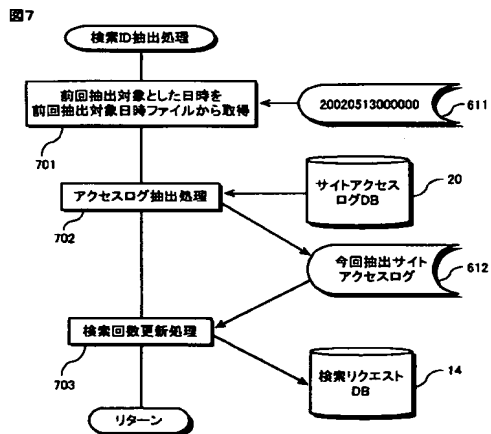
【図5】



【図6】



【図7】



【図8】

図8

(a)

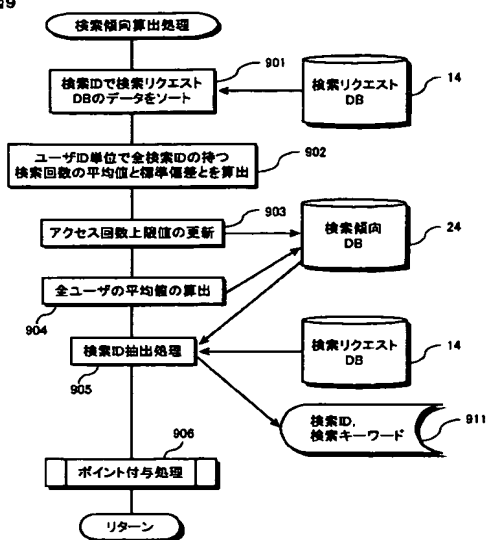
検索ID	連番	URL	アクセス日時
GAW0023514000001	001	aaa.bbb.org/	20020513141822
GAW0023514000001	002	www.research.bbb.com/~bs/	20020513141905
GAW0023514000001	003	www.ccc.com/	20020513142520
.....
GAW0023514000002	001
.....

(b)

ユーザID	アクセス回数上限値
0000000000	4.3
GAW0023514	3.5
JPS0088754	2.5
RDX0094812	1.2
FVE0088795	1.4
XJK0066478	5.6
.....

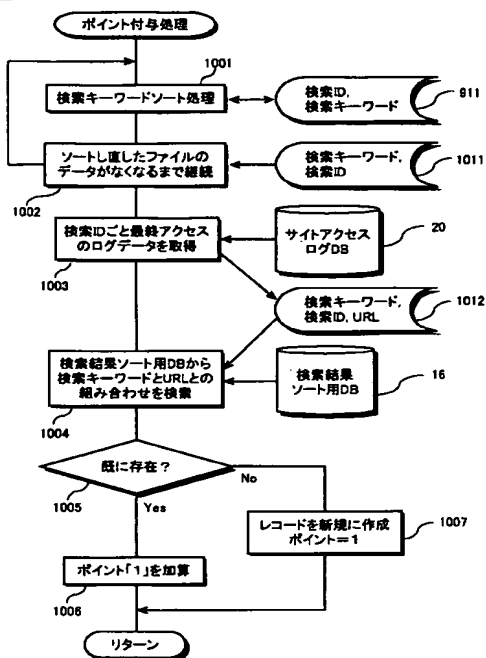
【図 9】

図9



【図 10】

図10



【図 11】

図11

